Consider the binary classification learning setting, where $\mathcal{Y} = \{0, 1\}$, some feature space $\mathcal{X}$, and the hypothesis space $\mathcal{H} = \{\pi : \mathcal{X} \rightarrow [0, 1] | \pi(\mathbf{x}) = s(\boldsymbol{\theta}^T \mathbf{x}), \boldsymbol{\theta} \in \Theta\}$ of logistic regression.

Exercise 1: Risk Minimizers for the Log-Loss

The first loss function of interest is the log-loss:

$$L(y, \pi(\mathbf{x})) = -y \log(\pi(\mathbf{x})) - (1 - y) \log(1 - \pi(\mathbf{x}))$$

(a) What is π_c^* (the optimal constant model in terms of the theoretical risk)?

(b) What is its risk?

(c) What is $\hat{\theta}$ (the optimal constant model in terms of the *empirical* risk)?

Exercise 2: Risk Minimizers for the Brier Score

Now consider the Brier score:

$$L(y, \pi(\mathbf{x})) = (\pi(\mathbf{x}) - y)^2$$

(a) What is π_c^* (the optimal constant model in terms of the theoretical risk)?

(b) What is its risk?

Loss	Risk minimizer	Bayes risk	Optimal constant model	Risk of optimal constant model
L2				
0/1				
Log				
Brier				